(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 January 2001 (04.01.2001)

PCT

(10) International Publication Number
WO 01/01305 A1

(51) International Patent Classification[6]: G06F 17/60, A61B 5/00

(21) International Application Number: PCT/US00/10727

(22) International Filing Date: 20 April 2000 (20.04.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/141,191	25 June 1999 (25.06.1999)	US
09/495,185	1 February 2000 (01.02.2000)	US
09/553,162	19 April 2000 (19.04.2000)	US

(71) Applicant: INTERNATIONAL DIAGNOSTIC TECHNOLOGY, INC. [US/US]; 121 Yancy Road, Madison, AL 35758 (US).

(72) Inventors: MADARASZ, Frank, L.; 121 Yancy Road, Madison, AL 35758 (US). INGUVA, Ramarao; 1200 Siniard Drive, Huntsville, AL 35803 (US). WYLY, James, K.; 18 Buckingham Drive, Bow, NH 03304 (US). MILELLI, Joseph; 684 Carnellon Court, Simi Valley, CA 93065 (US). KRIVOSHIK, David, P.; 113 Wertsville Road, Ringoes, NJ 08551-1108 (US).

(74) Agent: FERRONE, Diane; Gibbons, Del Deo, Dolan, Griffinger & Vecchione, One Riverfront Plaza, Newark, NJ 07102 (US).

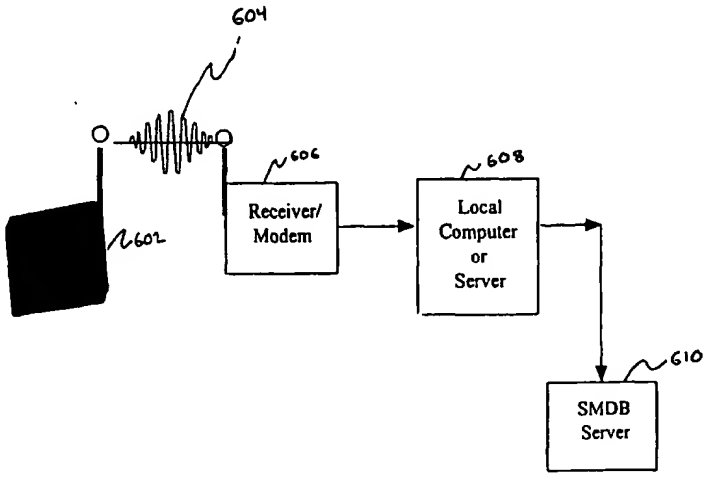
(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD AND SYSTEM FOR ACCESSING MEDICAL DATA



(57) Abstract: A system for accessing medical data comprises a computer system (figure 6, item 608, item 610) providing access to a collection of clinical medical data (figure 6, item 610) and a means for querying the collection of clinical medical data (figure 6, item 608) to determine a diagnosis and probability of successful diagnosis for a patient based upon assessment of the patient to obtain medical data.

WO 01/01305 A1

METHOD AND SYSTEM FOR ACCESSING MEDICAL DATA

CROSS REFERENCES TO RELATED APPLICATIONS

This application is a continuation in part of United States Patent Application Serial No. 09/495,185, entitled "Method And System For Accessing Medical Data" filed on February 1, 2000, and claims the benefit of United States Provisional Application Serial No. 60/117,509, filed on January 28, 1999 and United States Provisional Application Serial No. 60/120,309, filed on June 25, 1999.

BACKGROUND OF THE INVENTION

With the enormous amount of information medical information, making a decision on a factual basis is difficult. The collective experience of clinical practice forms the basis of virtually all health care decision making. Until now a readily accessible compilation of collective medical wisdom simply did not exist: virtually all information, which found its way to efficient dissemination, was of the research-results kind. While absolutely essential to support medical progress, this kind of data is essentially useless in the day to day practice of optimum medicine.

There is a need to provide ready access to large amounts of medical data either formerly unavailable or available in ways that were so cumbersome, or took so long, as to prohibit any real utility. This ready access can support the query prediction, diagnoses identification, etc., of the treating clinician.

SUMMARY OF THE INVENTION

The present invention is a system for accessing medical data comprises a computer system providing access to a collection of clinical medical data and a means for querying the collection of clinical medical data to determine a diagnosis and probability of successful diagnosis for a patient based upon assessment of the patient to obtain medical data. A method is also described.

BRIEF DESCRIPTION OF THE DRAWINGS

A more complete understanding of the present invention may be obtained from consideration of the following description in conjunction with the drawings in which:

FIG. 1 is a diagrammatic representation of the database architecture;

FIG. 2 is a graphical representation of probability distribution for a category of illnesses versus the range of illnesses in that category conditioned on only symptoms and any prior information, $P(\Psi \mid \mathbf{S}\, \mathbf{I})$ vs. Ψ;

FIG. 3 is a graphical representation of the most probable distribution of illnesses conditioned on physical examination, laboratory tests, diagoses ruled out, patient's medical history and any prior information, $P(\mathbf{D} \mid \Psi\, \mathbf{S'}\, \mathbf{I})$ vs. Ψ;

FIG. 4 is a graphical representation of probability distribution for various treatments conditioned only on the determined set of symptoms, the most probable illness (see Fig. 2.) and any other prior information $P(\mathbf{T} \mid \mathbf{S}_f \Psi_p \mathbf{I})$ vs. T;

FIG. 5 is a graphical of the most probable distribution of treatment conditioned on additional information, **d**, such as cost and insurance coverage, patient's preferences, additional information concerning the patient his medical history and medication reactions if not included in the prior information, **I**, $P(\mathbf{T} \mid \mathbf{d} \; \mathbf{S}_f \Psi_p \mathbf{I})$ vs. **T**; and,

FIG. 6 is a high level view of a representative implementation of the SMDB system.

DETAILED DESCRIPTION OF VARIOUS ILLUSTRATIVE EMBODIMENTS

While the present invention method and system for accessing medical data is particularly well suited for use with human patients and shall be so described, it is equally well suited for use with other species. The present invention method and system for accessing medical data offers significant advantages in the practice of veterinary medicine as well as is generally applicable to treatment of any living organism. Veterinary medical data can provide veterinarians as well as breeders with a powerful tool in improving the health of particular animals as well as the over all breed.

Clinical experience forms the proper basis to inform all decisions in health care, but it must be collected and carefully monitored. While the financial, social, and ethical issues must all be taken into account, these are areas already tracked and the subjects of lively discussion. If, however, a physician, faced with a presenting patient, wishes to know what the statistical experience of the last, say,

1000 patients so presenting has been, he is lost at sea. If that same physician, having made a well founded diagnosis, wishes to know the treatment choices made, and their relative outcomes of the last 1000 or so similarly diagnosed patients, he is similarly lost. But it is precisely these kinds of data on which the best medical decisions must be made. The development of the Smart Medical DataBase ("SMDB") provides a resolution of this dilemma.

The SMDB will provide ready access to huge amounts of data either formerly unavailable, or available in ways that were so cumbersome, or took so long, as to prohibit any real utility. The very first advantages are the availability of the data itself, and quick access in a clinically meaningful manner, made possible by sophisticated sorting algorithms.

After an initial period, patterns of use, and feedback from users, will inform future SMDB expansions and modifications to improve its utility. For example, often-requested series of questions about a disease complex could be placed under a single command. In addition to such changes, adaptive responses of the system become part of the system itself.

At every stage the data that is used, and how it is used, is solely at the discretion of the treating physician. The longstanding and widespread failure of so called expert systems until now, such as those which read EEG's and EKG's, are essentially sophisticated sorting programs, simply matching values of multidimensional parameters with suggested diagnoses.

A strength of the SMDB system is that when more sophisticated "intelligent" functions are introduced, they are already in the statistical test bed in which they may be evaluated for effectiveness. A new drug is evaluated by clinical trials, undertaken after a very small research sampling of effect. The clinical trials themselves, while wider in scope, still remain small relative to eventual field use. After long-term use in the field, a much larger base of data becomes available about efficacy, side effects, etc.

The SMDB however, in any of its evolved stages, can be said to be self-testing without clinical risk. As data on the reliability and accuracy of a certain SMDB function accumulate, then it become increasingly relied upon by the user community. The ever-expanding database is always available in real time and constantly up to date.

The decision-making capabilities of the Medical Protocol and its associated database structure, to be described in detail below, come under the general heading of AI. What is proposed for the SMDB can be classified under specific components of AI: Logical AI, Inference, and Learning from Experience. Learning from Experience is the area of "smartness" in the SMDB. The tools most commonly employed to implement this component are connectionism, neural networks, semiotics, and fuzzy logic. The present invention add the use of Bayesian Statistics. The Bayesian Method specifically focuses on a predictive/learning capability. Moreover, its statistics are very well developed and mature compared to the tools presently being used in AI. The present invention is

not limited exclusively to the Bayesian Method for the Learning from Experience component of the SMDB, but consider all viable tools used in AI and those to be developed. However, it is a powerful method, uniquely different from the tools presently employed in AI.

While clinicians, the care providers, are a primary target user of the present invention, insurers and HMO's, the payers, can add their already formidable financial data to the relevant subset of corresponding clinical information and achieve a truly effective tool for clinically informed business management. Similarly, the managers of organizations, as well as public policy makers, can use the same comprehensive database to make policy decisions genuinely grounded in reality.

The following symbols and abbreviations are used in this application:

C/C – Chief Complaint
C/O – Complaining Of
Dx – Diagnosis
{ ... } – Set, or Collection of, ... ; as in {Dx}
F/U – Follow Up
Glu – Glucose
Hx – History
NAD – No Apparent Distress
PE – Physical Exam
PT – Patient
R/O – Rule Out
Sx – Symptom—Something patient brings you
Sign – Something elicited from patient by you
URI – Upper Respiratory Infection
WBL – White Blood Cell count

WNL – With Normal Limits

Referring to Fig. 1: there is shown the static organizational structure (architecture) of an embodiment of the SMDB. The SMDB has three major divisions. The two represented by circles, the Diagnostic segment 102 and Treatment segment 104 are primarily clinical in origin. The triangular portion 106 consists of that data driven by and collected about costs and payments. These distinctions reflect the major source of the data in each segment: the multiple overlapping diagrams indicates that whatever the source, portions of each division will ultimately be of interest to everyone. The payer is interested in relative treatment costs and their relation to efficacy, while the MD provider must increasingly be aware of costs. For instance, the solid circle 108 represents that small portion of data from all three sources to which the patient has direct access. The striped portion 110 might be a similar portion accessible to the payer: basic clinical and cost information without compromising patient confidentiality. There are as many configurations as there are users.

In a more succinct mathematical approach, the SMDB is comprised of three major parts: The **D** (**D**iagnostic) part 102; the **T** (**T**reatment—**O**utcome) part 104; the **M** (**MIS**—Management Information Systems) part 106. The Venn diagram picture shown in FIG. 1 is one where all three parts overlap, resulting in seven (7) distinct pieces: the **D, T, M, DT, DM, TM**, and **DTM.**

The **D** part 102 subdivides into the three-tuple of signs, symptoms, test results **(SST)** mapped into the diagnostic set **Dx**—this is most easily and

productively seen as the ICD codes. (The symbol "**D**" represents one of three main parts of the database itself, while the "**Dx**" represents a particular set, the set of individual diagnoses, within that bigger part **D**.) The ICD is a worldwide classification scheme, mostly descriptive but etiologic when it can be, that is used for almost everything, including billing. The world uses ICD-10, while the USA alone is still using ICD-9. Strictly, the mapping into is incorrect, since there are three-tuple images without images in **Dx**, and members of **Dx** without inverse images in the SST. **D:Dx:(S,S,T)→Dx**

The **T** part 104 contains maps that take **Dx** (and also separately a part of SST) into the treatment set, **Tx**, (The same kind of relationship holds for the **T** and **Tx** distinction as for the **D and Dx** sets above: **T** is a major segment of the data base, and it contains, amongst many other things, the set of treatment protocols, {**Tx**}) and then the **T** set of these maps to an **O**utcome set. The **O**utcome set is an n-dimensional array with clinically relevant measures which may be in terms unique to the **O**utcome set, but which may contain also members of the SST set: e.g., normalization of white count; five year survival; return to pain-free state; resolution of fever; normalization of vital signs, etc. The direct SST to treatment maps describe symptomatic treatment with no ICD identity, admittedly a somewhat rare occurrence. **T:Dx→Tx→O**utcome

The **M** part 106 of course has members both in the **D** and **T** sets (but not all), while also having elements unique to itself. These latter contain all the usual tools and measures typical of **MIS** systems, and a few unique to the SMDB.

The idea associated with the seven pieces described above is being able to have selective availability. For the majority of physicians, access to part **D** 102 and part **T** 104 will be dominant, while those managing their own practices might have limited availability to part **M** 106. The harmonious union of clinical management plus financial information comes from the fact that the billing aspect of practice is where the two come together. An HMO or other insurer will be most interested in the **M** section 106. Hospitals will take an interest in all three pieces, and it is anticipated that most of Quality Assurance and Utilization Review (QA and UR) functions will be subsumed and much more efficiently accomplished using the SMDB organization. Notice that the **M** part 106 may also contain, in addition to data, sophisticated tools now used by the industry.

It is anticipated that most of what is needed for all three parts is already collected under the aegis of clinical record keeping and the billing function. Any necessary changes will be almost resource-invisible in nature, mainly consisting in changes to medical forms (including electronic), and how data is organized and made selectively available (software changes).

A brief example follows to outline the basic aspects of the Bayesian approach. The number for the probability for the disease is derived using standard statistical tools as well as Bayes theorem. Let **H** be the hypothesis whose validity is being examined. (Please note that in what follows all boldfaced symbols imply vector quantities, i.e. a possible n-tuple of symptoms, tests, diseases, etc.) In other words,

H(x): The symptoms and test results imply disease(s) **x**.

Since the set of symptoms and test results are never exhaustive, the validity of **H** in general has to be statistical based on evidence. The confidence in the validity of **H(x)** is summarized by the quantity P(**H(x)**|**DI**) defined by

P(**H(x)**|**DI**) = Probability Density(confidence) that **H(x)** is true.

Conditioned on the evidence **D** and prior information **I**,

D: Set of all symptoms and test results,

and as usual $0 < P(\mathbf{H(x)}|\mathbf{DI}) < 1$.

P(**H(x)**|**DI**) =1 implies 100% confidence in **H(x)**

P(**H(x)**|**DI**) = 0 implies **H(x)** is ruled out

The estimation of P(**H(x)**|**DI**) is done using Bayes theorem:

$$P(\mathbf{H(x)}|\mathbf{DI}) = P(\mathbf{D}|\mathbf{H(x)I})\, P(\mathbf{H(x)}|\mathbf{I}) / P(\mathbf{D}|\mathbf{I})$$

where

P(**H(x)**|**I**) = Prior probability density that **H**(x) is true based on previous symptoms and tests on patients other than the one currently under examination,

P(**D**|**H(x)I**) = Likelihood of **H(x)** based purely on current evidence. Confidence based on current symptoms and tests,

and

P(**D** | **I**) = Probability Density for **D**.

The medical database system offers an estimation of $P(\mathbf{H}(x)|\mathbf{I})$ using all the information available in its database. It will also estimate the likelihood based on current evidence. Based on the database it can also recommend what symptoms to look for and what tests to order based on previous case histories.

The query responses are a function of the number of symptoms that matched with the list of known symptoms for a particular disease state, the history of the patient based on this match of the symptoms, and commonality characteristic with symptoms in other illnesses. Unless requested this response uses the entire sample of patient data on the database. The form also has options to choose specifics that would restrict the sample by cross matching to selected demographic and medical variables. The software to choose a sub-sample of the entire data set and come up with a new real time response would then use this later option.

The software would also query the doctor if he wants to see an additional list of symptoms that might increase the probability of a particular diagnosis. To confirm the diagnosis the doctor in consultation with the patient might look for additional symptoms over a stipulated time period, or order tests that might reveal additional symptoms. He would then repeat, if necessary, the above procedure to come to a firm decision as to the nature of the illness. At this stage we assume that the doctor in consultation with the patient and through his software and additional tests, has come to a decision regarding the illness.

Regarding the choice of treatments, the system presents the option of running the treatments options menu. Depending on the menu choices, the

software gives details about the samples of patients and the response to a particular treatment as a function of time. It will present a number of statistical processing results depending on the queries posed up by the doctor (chosen out of a list or entered by the doctor) such as probability of cure versus probability of failure of treatment. This is done interactively as a function of time. The software will give a set of alternatives as a guide to proceed. It will also provide a list of contacts for consultation should it prove necessary. It will also search additional databases should further processing be necessary. If the doctor incorporates the patient treatment and response into the software, it can in real time compare/match with similar patient scenarios and give useful pieces of information such as recommendations for midcourse change of medicines/treatment etc.,

The people authorized by the patient and doctor such as the on-call physician will put the information on a web site for use. The doctor using the same software can use a utility to post a question on the bulletin board, protecting the patient's identity, to which other people on the network can respond.

SMDB updates occur in real time, allowing analysis and decision making at the time and place of physician contact. In addition, correlations can be had sorting for patients with constellations of illnesses and conditions, and all the relevant probabilities of diagnoses, treatments, and outcomes derived for any specialized situation. In this way the vast amount of accumulated clinical data may be queried for the maximum it can inform about a particular patient.

Two functions of the SMDB which will have increasing importance in the future will be the telemetering of patient data via phone/cable lines to physicians' offices and hospitals, and the transmission of lab results and imaging data for the best expert interpretation. The former will be of use in both crisis management (e.g. asthma attacks) and maintenance (e.g. blood pressure, glucose, or pulmonary function monitoring).

Further along in a case, the SMDB will allow an informed analysis of failed treatments and treatment alternatives.

Lastly, such a massive integration of clinical data will allow relatively effortless retrospective studies of diagnostic accuracy, treatment efficacy, and safety. It will permit, within confidentiality legislation and guidelines, the convenient access of individual patient histories and data by those people and agencies entitled to that data.

The Bayesian Statistical Approach

The Bayesian approach gives the posterior probability density as a function of the prior density functions, while a utility function is used to order preferences. These things are defined carefully in terms of our envisioned structure and possible queries. Decisions relying on the Bayesian approach use the following:

A space of possible actions by the decider – the clinician.

- A space of possible patient states, diagnoses for example.
- A collection of possible experiments, or tests, or findings.
- A space of possible outcomes for these tests, experiments.

- A utility function defined over all the possible consequences which defines outcome preferences.
- The above are then used to "condition" the prior probability density function as more results (information) become available, yielding the posterior probability.

While SMDB can be implemented with a centrally located data base accessed over a computer network such as the internet using, utilizing established communication protocols and browser type interfaces, the present invention is also equally well suited for a distributed architecture implementation.

The effective data compression provided by Bayesian Statistics lends itself to a distributed architecture implementation. For example, millions of patient records are reduced to sets of conditional probabilities, which only require 4-digit resolution to be effective. Consider the following: a billion patient records can be distilled into a table containing, perhaps 10,000 sets of conditional probabilities. As the size of the database increases, the table values may change; however, the size of the table is unchanged.

In one embodiment of the SMDB, the collection of patient records is hosted on a network accessible, centralized system. Each user is provided with software enabled to run on an individual computer or over a local network. The software consists of the user interfaces; the algorithms described herein and the tabular database of conditional probabilities and interactive diagnostic query responses. Thus, the physician will be able to utilize the database in an efficient mode.

There are several methods that can be used to update the database. One method employs "push" type technology, wherein the central database sends, daily updates of the tables as well as other relevant information to the user. Another method employs downloading updates, such as when a patient's records are uploaded to the SMDB central database. Either method is efficient as the updates would be minimal in volume and typically only require a few seconds to transfer.

However, the simpler option of accessing the centralized SMDB, with a simple browser, and performing the patient data upload and diagnostic query/response mode is useful when physicians are away from an office environment and can access the system using cell phone, personal digital assistant, such as a palm pilot, or lap top computers.

The following is an simple exemplary embodiment of the SMDB and the application of Bayesian Methodology to a Clinical Case.

Statement of Illness and Background Information on Patient

Mr. Sam Jones is a 37-year-old married white male who presents with multiple upper respiratory complaints. He is new to this clinic secondary to his having recently changed employers, and hence insurance carriers. He currently works as a janitor at a local school. Chief Complaint: 'I have this cough, and my throat is very sore.' Complains of night sweats, waking the patient two to three times a night for the last three nights. Complains of cough, non-productive at first, now with mucopurulent discharge, occasionally blood tinged, times five weeks.

Construct Initial Probability Distribution

$$P(\Psi \mid \mathbf{S}\,\mathbf{I}) =$$

Conditional Probability for category of Illness - Ψ (Vector)
Conditioned on Symptoms Vector **S** and Any Other Prior Information
Concerning the Patient and Their Medical History

Any illness $\Psi_1, \Psi_2, \Psi_3, \ldots \subset \Psi$ and Ψ_p is likely. The Ψ_i's are a range of illnesses that correspond to the stated symptoms given in **S**, and Ψ_p is the final outcome—a specific illness with an associated probability of being true—of the analysis. The result of this construct would typically look as that displayed in Fig. 2.

Medical Examination

Physical Examination

1. Vital signs
    a. Blood pressure: 124/70
    b. Pulse: 85
    c. Respiration: 20

d. Temperature: 101 degrees Fahrenheit

2. General description: the patient is a 37 year old married white male who looks his stated age; he is pleasant, appears well nourished, and seems in an overall good state of health.

3. Skin: warm and dry; turgor adequate; color normal. There is no icterus, purpura, rash, or unusual pigmentation noted. Hair normal in appearance, distribution, texture.

4. Lymph nodes: no cervical, supraclavicular, axillary, epitrochlear, or inguinal adenopathy.

5. HEENT (Head, Ears, Eyes, Nose, Throat):

a. Head: normocephalic and atraumatic; no lesions noted.

b. Eyes: cornea without lesions, conjunctiva clear, sclera white. Pupils are equal, 3mm in diameter, round, reactive to light and accomodation. Extraocular movements within normal limits without nystagmus or strabismus. Fundii are benign. Disks well delineated. There are no hemorrhages or exudates. Visual acuity is 20/20 bilaterally, and visual fields are within normal limits to confrontation.

c. Ears: normal in appearance. Auditory canals clean and without lesions. Tympanic membranes intact. Hearing adequate.

d. Nose: septum within normal limits and without deviation. Nasal mucosa pink and without abnormal discharge. No nasal polyps or other lesions. Frontal and maxillary sinuses nontender.

e. Mouth and throat: lips without cyanosis or pallor. Buccal mucosa normal in appearance. Teeth in good condition. Tongue without lesions or tremor, protrudes midline. Pharyngeal mucosa is erythematous and without other lesions, exudates, or evidence if inflammation. Gag reflex intact.

6. Neck: neck is supple, full range of motion. No evidence of tracheal deviation, jugular venous distension, or lymphadenopathy. Carotid pulses are 2+, equal bilaterally, and without bruits. Carotid upstroke is within normal limits. Thyroid normal in size, palpation reveals no nodules or masses.

7. Back: spinal curvature is normal; no scoliosis, kyphosis, or tenderness. Full range of motion present.

8. Chest: thorax is symmetric. Full expansion bilaterally. AP diameter is within normal limits.

9. Lungs: fremitus is equal bilaterally. Lung fields resonant throughout. Breath sounds and voice sounds normal. There are no rales or ronchi, but some end-expiratory wheezes throughout, more prominent in the bases bilaterally.

10. Heart: palpation reveals no heaves or thrills. The PMI (point of maximum impulse) is medial to the midclavicular line, fourth intercostal space. Auscultation reveals S1, S2, of normal intensity. There are no S3, S4, rubs, clicks, or other abnormal heart sounds. Heart rate is 70 BPM and rhythm is regular.

11. Breasts: breasts are symmetric and of normal contour. Skin is of normal color and appearance; there is no edema, ulceration or erythema. Nipples are of normal size and shape; there is no nipple retraction, ulceration or discharge. Palpation does not reveal any tenderness or masses.

12. Abdomen: normal size and contour. There are no capillary dilatations, skin lesions, or surgical scars. Auscultation reveals normative bowel sounds and no abdominal bruits. Palpation reveals no abdominal tenderness, guarding, or masses. Liver edge is felt approximately 1 inch below the right coastal margin; it is firm, sharp, and smooth. The liver percuses to approximately 8 to 10 cm. total span. The spleen is not palpable.

13. Rectal exam: no external anal lesions. Sphincter tone normal. No internal or external hemorrhoids. Rectal mucosa appears normal, with no nodules or masses present. Stool is brown and negative for occult blood.

14. Genitalia: inspection reveals normal distribution of pubic hair. No lesions or discharges. No external lesions. Testes are descended, nontender, of normal size, without nodules or masses.

15. Inguinal area: no lymphadenopathy noted. Femoral pulses are 2+ and equal bilaterally. Auscultation reveals no femoral bruits.

16. Extremities: there is no clubbing, cyanosis, or edema. Brachial, radial, popliteal, dorsalis pedis, and posterior tabialis pulses are 2+ and equal bilaterally. Musculoskeletal exam reveals no joint deformities and full range of motion. There is no bone, joint, or muscle tenderness noted.

17. Neurologic: patient is alert and oriented to time, person, and place. Cranial nerves II to XII are within normal limits. Speech, memory, and expression are within normal limits. Muscle strength is 5/5 in both upper and lower extremities. There is no muscle atrophy or involuntary movement noted. Testing of cerebella function reveals normal gait, negative Romberg test, and good coordination in finger-to-nose, heel-to-shin, and alternate motion testing. Sensory is intact to light touch, pain, and vibratory stimuli. There are no focal motor/sensory deficits. Deep tendon reflexes are 2+ and equal bilaterally.

Chosen Tests Include:

1. Throat culture and sensitivity (including strep): rapid strep test (poor but widely used) is negative. 24 and 48 hour cultures are positive for streptococcus sensitive to a wide range of older antibiotics. Mycoplasma culture will take a week or more to return, somewhat more difficult to do reliably, and is not chosen.
2. Chest X-ray, PA and LAT (back to front and side views). The pictures show very light patchy bilateral infiltrates, without consolidation.
3. A PPD (TB test) is chosen, and planted. A mycobacteria culture is not chosen (possibly a mistake) at this time.

The Diagnoses to be Ruled Out Include:

1. Strep infection
2. Pneumonia
3. Residual mycoplasma infection

4. Mycobacteria infection (tuberculosis)

Construct Data Set (Vector) D and New Symptoms Vector S'

The data set is made up of the information contained in the seventeen (17) steps of the Physical Examination, that obtained from the three (3) Laboratory Tests and from the four (4) Diagnoses Ruled Out:

$$\mathbf{D} = \{(D_1, D_2, \ldots D_{17}\text{—Physical Exam}), (D_{18}, D_{19}, D_{20}\text{—Lab Tests}),$$

$$(D_{21}, D_{22}, D_{23}, D_{24}\text{—Ruled Out})\}$$

In addition, the physician and/or the patient may have identified one or more symptoms not initially reported. The symptoms vector must be updated:

$$\mathbf{S'} = \mathbf{S} + \text{Additional Symptoms Observed.}$$

Construct a New Probability Distribution for the Illness Ψ Using D, S' and Bayes' Theorem

$$P(\Psi \mid \mathbf{D}\, \mathbf{S'}\, \mathbf{I}) =$$

Conditional Probability for Reduced Range of Illnesses - Ψ (Vector)
Conditioned on New Data D, Updated Symptoms Vector S' and Any Other Prior Information
Concerning the Patient and Their Medical History

$$P(\Psi \mid \mathbf{D}\, \mathbf{S'}\, \mathbf{I}) = [P(\mathbf{D} \mid \Psi\, \mathbf{S'}\, \mathbf{I})\, P(\Psi \mid \mathbf{S'}\, \mathbf{I})] \,/\, P(\mathbf{D} \mid \mathbf{S'}\, \mathbf{I}),$$

where

$P(\Psi \mid \mathbf{S'}\, \mathbf{I})$ = Prior probability density that Ψ is true based on previous symptoms and tests on patients other than the one currently under examination,

$P(\mathbf{D} \mid \Psi\, \mathbf{S'}\, \mathbf{I})$ = Likelihood of Ψ based purely on current evidence.

Confidence based on current symptoms and tests,

and

$P(\mathbf{D} \mid \mathbf{S'}\, \mathbf{I})$ = Probability Density for **D**.

The new distribution is much narrower in the range of illnesses it spans and peaks about the most probable illness, Ψ_p as illustrated in Fig. 3. At this point, the physician, based on other past case experiences with similar symptoms and/or professional instinct, has an option of selecting the most probable illness, Ψ_p, to treat or one of a small number of other illnesses within a nominal 0.99 probability "threshold" as indicated in Fig. 3.

A Presumptive Diagnoses of possible strep infection, possible mycoplasma infection is made. Less likely is TB. The patient will be treated as an outpatient.

The SMDB is accessed for presenting symptoms and signs. It is found that in the last 30 days there have been 19 cases of students or staff of that school diagnosed with mycoplasma infection. This makes the diagnoses now overwhelmingly likely.

All have responded to a course of erythromycin (not universally true for this infection). Inhaled bronchodilators, and inhales steroids.

The Patient Treatment Plan

The SMDB is designed to be an effective tool in helping to select a course of treatment for the patient. Two in puts are needed: the most probable illness, Ψ_p, and a final determined set (vector) of symptoms, $\mathbf{S}_f$, which may be identical to **S'**.

When this information is put into the SMDB it returns a probability distribution of treatment, **T**,

$$P(\mathbf{T} \mid S_f \Psi_p \mathbf{I}) =$$

$$\left\{\begin{array}{l}\text{Conditional Probability of Treatment(s), } \mathbf{T} \\ \text{Conditioned on Symptoms Vector } S_f \text{ and Any Other Prior Information } \mathbf{I} \\ \text{Concerning the Patient, His Medical History Including Medication} \\ \text{Information}\end{array}\right.$$

As illustrated in Fig. 4, the return may be a suggested board range of treatments, $T_1, T_2, \ldots$.

Tailoring the Treatment Plan

The treatment plan may be tailored to conform to additional information. First a new data set (vector), **d**, is constructed. This data set will contain information such as simplicity of treatment, cost and insurance coverage, patient's preferences, additional information concerning the patient his medical history and medication reactions if not included in the prior information, **I**, etc., etc. From this information the SMDB will employ the Bayesian Methodology and construct the most probable distribution of treatments.

$$P(\mathbf{T} \mid \mathbf{d}\, S_f \Psi_p \mathbf{I}) = [P(\mathbf{d} \mid \mathbf{T}\, S_f \Psi_p \mathbf{I})\, P(\mathbf{T} \mid S_f \Psi_p \mathbf{I})] / P(\mathbf{d} \mid S_f \mathbf{I}),$$

where

$P(\mathbf{T} \mid S_f \Psi_p \mathbf{I})$ = Prior probability density that **T** is true based on the determined illness, Ψ_f, the symptoms, S_f, associated with Ψ_f and other prior information, **I**.

$P(\mathbf{d} \mid \mathbf{T}\ \mathbf{S}_f\ \mathbf{I})$ = Likelihood of **T** based on no directly knowledge of Ψ_p.

and

$P(\mathbf{d} \mid \mathbf{S}_f\ \mathbf{I})$ = Probability Density for **d**.

The new distribution is much narrower in the range of treatments it spans and peaks about the most probable illness, T_p as illustrated in Fig. 5. At this point, the physician, based on other past case experiences with similar situations and/or professional instinct, has an option of selecting the most probable treatment, T_p, or one of several other treatments that nominally fall within 0.99 probability "threshold" as indicated in Fig. 5.

A course of erythromycin (both because of the above information, and to nail the last doubt about a possible strep infection, and helping to prophylax against the latter).

- A two week course of an inhaled bronchodilator.
- A two week course of an inhaled steroid.
- Return in two days to have PPD read.
- Return in two weeks for follow up and to schedule an initial intake to the service.

The PPD is negative in two days: no action taken. Patient reports good symptomatic relief with the inhaled medication.

Two-week follow up shows a satisfactory resolution of symptoms: productive cough is absent times 8 days; patient is afebrile. Lungs are now clear. No follow up chest X-ray is needed. Oropharynx is now clear with no erythema.

Update/Upgrade SMDB

This patient's medical event now become part of the SMDB, and hence part of the knowledge informing subsequent decisions for other uses of the SMDB.

While this case does not involve life and death issues, as do cancer diagnosis and treatment, it is sufficient to illustrate the utility of the present invention SMDB. Unless the treating team was directly involved in the other cases from the school, or had prescient knowledge about them, this confirmatory data would not be available. While in this case the resulting benefits included higher confidences in diagnosis and treatment, increased treatment efficacy, lower resource consumption all around, (all worthwhile goals), in most cases in the management of serious chronic illness, these advantages will accrue a thousand fold more. In addition, the issues of quality of life, morbidity and mortality, resource availability, access to treatment, are all vitally important areas, which will see significant benefits from the SMDB function.

Moreover, a portion of the SMDB subsumes the typical medical record and administrative record functions. This alone will not only enhance those roles, but also reduce costs involved with them.

Referring to Fig. 6 there is shown a representative implementation of the SMDB system. A personal digital assistant 602, used by the care provider,

contains client software and intelligent agents for providing user interfaces and communicates by a wireless link 604 to a wireless communication device 606. The remote communication device 606 is coupled to a local computer system or server 608. The local computer system or server 608 accesses the SMDB server 610, through the Internet, dedicated network, or other suitable communication links.

Numerous modifications and alternative embodiments of the invention will be apparent to those skilled in the art in view of the foregoing description. Application to systems of other living organisms as well as electro/mechanical systems having non specific diagnosis for failure or problems are within the scope of the invention. Accordingly, this description is to be construed as illustrative only and is for the purpose of teaching those skilled in the art the best mode of carrying out the invention. Details of the structure may be varied substantially without departing from the spirit of the invention and the exclusive use of all modifications which come within the scope of the appended claim is reserved.

WHAT IS CLAIMED:

1. A method for accessing medical data comprising the steps of:

assessing a patient to obtain medical data;

accessing a collection of clinical medical data;

querying the collection of clinical medical data; and

determining a diagnosis and corresponding probability of successful diagnosis for the patient;

wherein said corresponding probability of successful diagnosis is defined by $P(\mathbf{H}(\mathbf{x})|\mathbf{DI})$, probability density that said diagnosis, $\mathbf{H}(\mathbf{x})$, is true, said diagnosis $\mathbf{H}(\mathbf{x})$ is a function of symptoms and test results $\mathbf{x}$, $\mathbf{D}$ represents evidence and $\mathbf{I}$ represents prior information.

2. The method as recited in claim 1 wherein the step of accessing the collection of clinical medical data utilizes a computer network.

3. The method as recited in claim 2 wherein said computer network is a global computer network.

4. The method as recited in claim 1 wherein the collection of clinical medical data is distributed.

5. The method as recited in claim 1 wherein $P(\mathbf{H}(\mathbf{x})|\mathbf{DI})$ is estimated using Bayes theorem.

6. The method as recited in claim 1 wherein $P(\mathbf{H}(x)|\mathbf{DI})$ is estimated as $P(\mathbf{D}|\mathbf{H}(x)\mathbf{I})\ P(\mathbf{H}(x)|\mathbf{I}) / P(\mathbf{D}|\mathbf{I})$, where $P(\mathbf{H}(x)|\mathbf{I})$ is prior probability density that $\mathbf{H}(x)$ is true, $P(\mathbf{D}|\mathbf{H}(x)\mathbf{I})$ is likelihood of $\mathbf{H}(x)$ based on current evidence and $P(\mathbf{D} \mid \mathbf{I})$ is probability density for **D**.

7. The method as recited in claim 1 wherein the patient is a human.

8. The method as recited in claim 1 wherein the patient is an animal.

9. A system for accessing medical data comprising:

a computer system providing access to a collection of clinical medical data; and

a means for querying the collection of clinical medical data to determine a diagnosis and corresponding probability of successful diagnosis for a patient based upon assessment of the patient to obtain medical data;

wherein said corresponding probability of successful diagnosis is defined by $P(\mathbf{H}(x)|\mathbf{DI})$, probability density that said diagnosis, $\mathbf{H}(x)$, is true, said diagnosis $\mathbf{H}(x)$ is a function of symptoms and test results **x**, **D** represents evidence and **I** represents prior information.

10. The system as recited in claim 9 wherein access to the collection of clinical medical data utilizes a computer network.

11. The system as recited in claim 10 wherein said computer network is a global computer network.

12. The system as recited in claim 9 wherein the collection of clinical medical data is distributed.

13. The system as recited in claim 9 wherein $P(\mathbf{H}(x)|\mathbf{DI})$ is estimated using Bayes theorem.

14. The system as recited in claim 9 wherein $P(\mathbf{H}(x)|\mathbf{DI})$ is estimated as $P(\mathbf{D}|\mathbf{H}(x)\mathbf{I})\ P(\mathbf{H}(x)|\mathbf{I}) / P(\mathbf{D}|\mathbf{I})$, where $P(\mathbf{H}(x)|\mathbf{I})$ is prior probability density that $\mathbf{H}(x)$ is true, $P(\mathbf{D}|\mathbf{H}(x)\mathbf{I})$ is likelihood of $\mathbf{H}(x)$ based on current evidence and $P(\mathbf{D} \mid \mathbf{I})$ is probability density for $\mathbf{D}$.

15. The system as recited in claim 9 wherein the patient is a human.

16. The system as recited in claim 9 wherein the patient is an animal.

17. The system as recited in claim 9 wherein the computer is a personal digital assistant.

18. The system as recited in claim 9 wherein the means for querying the collection of clinical medical data utilizes a wireless communication interface.

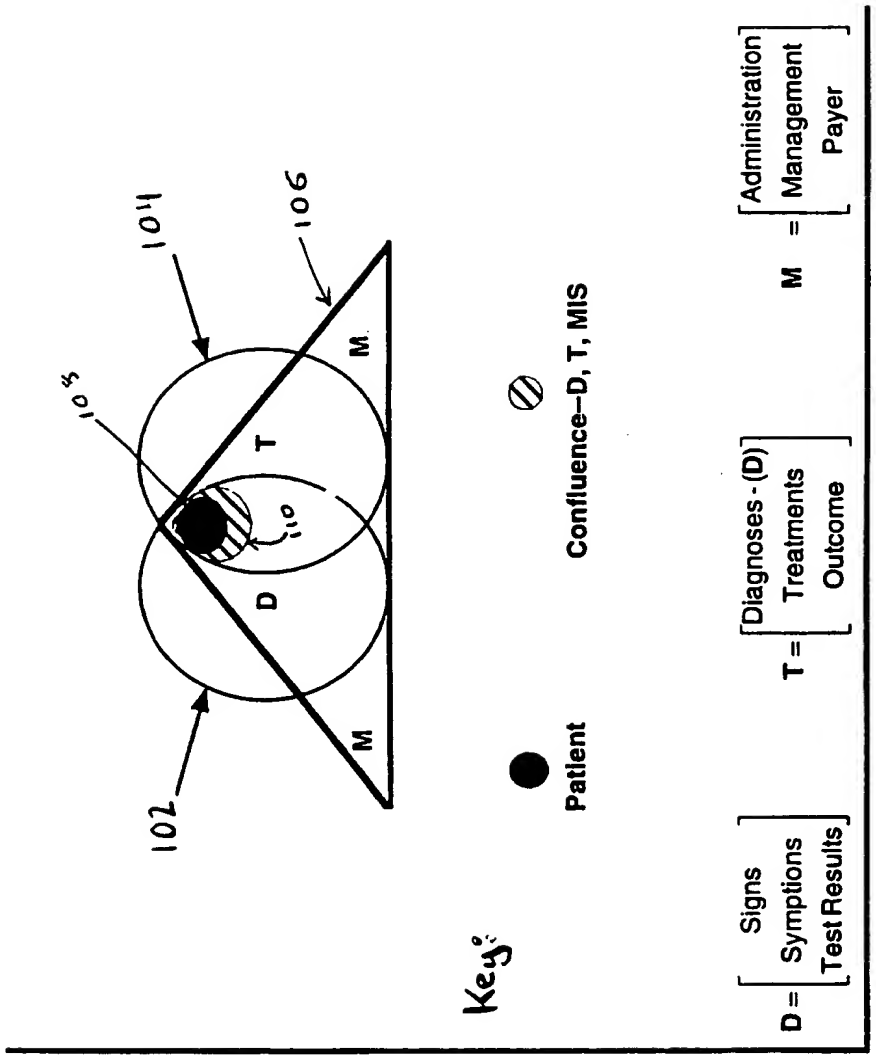

1011
106
105
102
110
T
D
M
M
Key:
Patient
Confluence–D, T, MIS
D= [Signs, Symptions, Test Results]
T= [Diagnoses - (D), Treatments, Outcome]
M = [Administration, Management, Payer]

Fig. 1

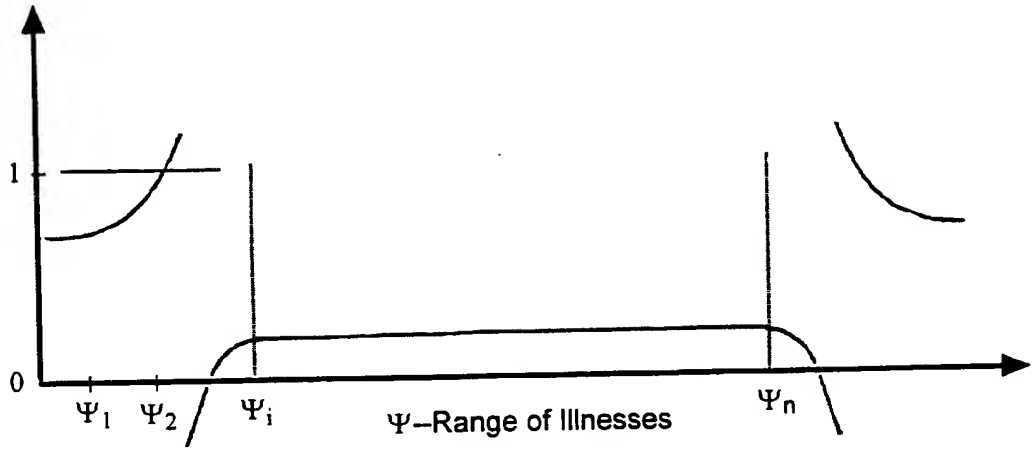

1
0
Ψ_1
Ψ_2
Ψ_i
Ψ_n
Ψ–Range of Illnesses

Fig. 2

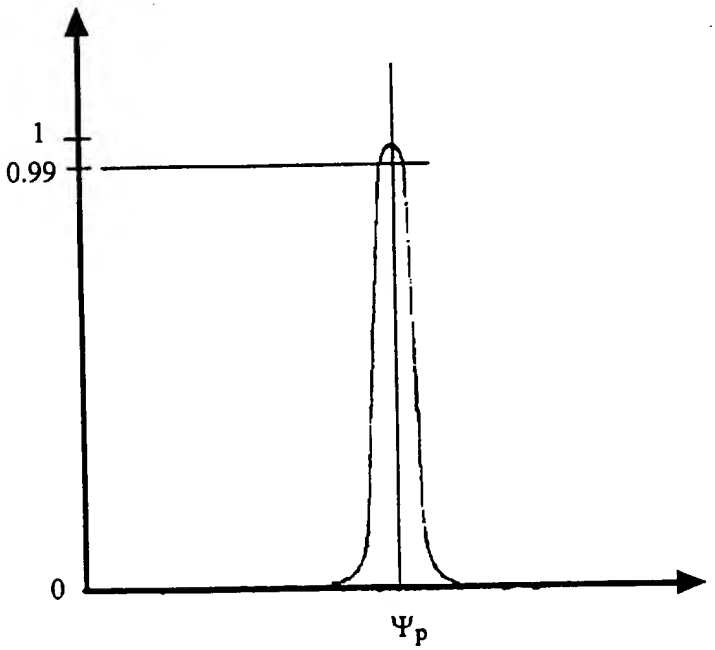

1
0.99
0
Ψ_p
Ψ--Range of Illnesses

Fig. 3

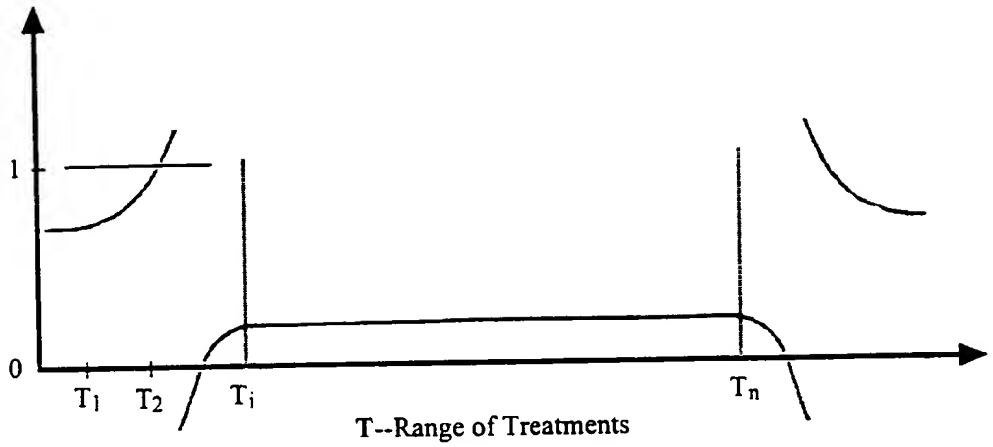

1
0
T_1
T_2
T_i
T_n
T--Range of Treatments

Fig. 4

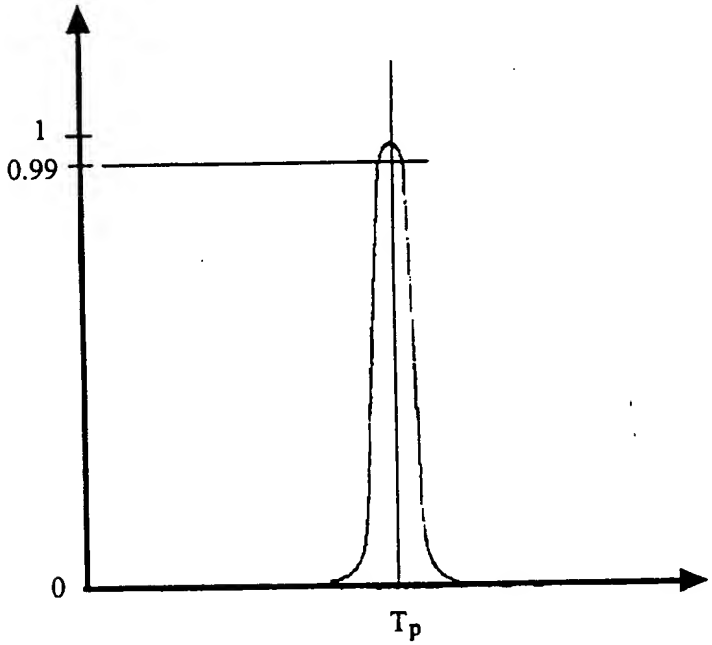

1
0.99
0
T_p
T--Range of Treatments

Fig. 5

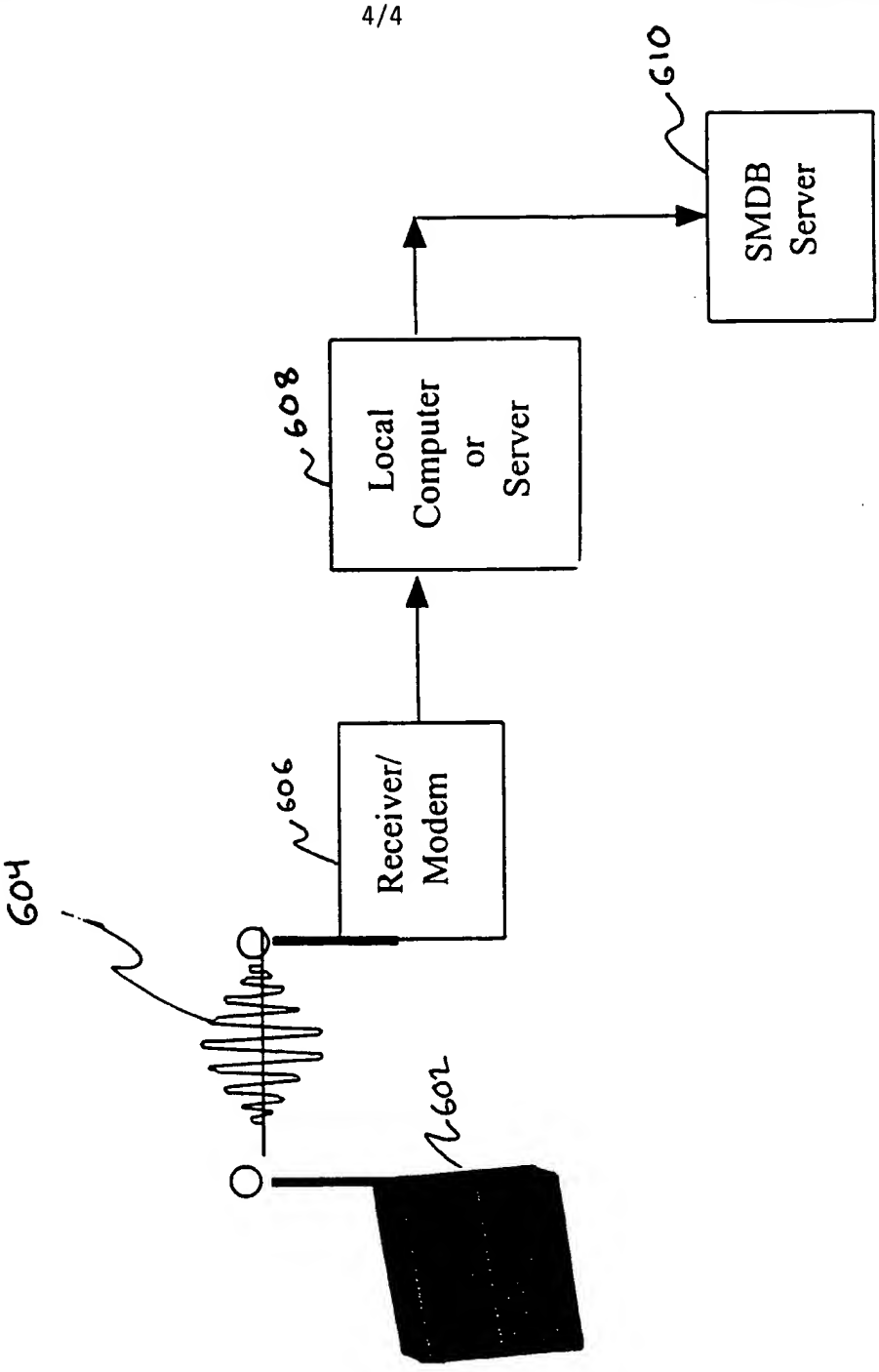

610
SMDB Server
608
Local Computer or Server
606
Receiver/ Modem
604
602

Fig. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/10727

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : G06F 17/60; A61B 5/00
US CL : 705/2, 3; 600/300

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 705/2, 3; 600/300

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EAST

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,133,046 A (KAPLAN) 21 July 1992 (21.07.1992), abstract.	1-18
A, P	US 5,935,060 A (ILIFF) 10 August 1999 (10.08.1999), column 3, lines 51-65, column 4, lines 1-11, column 7, lines 20-32, column 8, lines 39-52, column 11, lines 18-24, figure 1.	1-18
A, E	US 6,076,083 A (BAKER) 13 June 2000 (13.06.2000), abstract.	1-18
A, P	US 5,911,132 A (SLOANE) 08 June 1999 (08.06.1999), abstract.	1-18
A, P	US 6,024,705 A (SCHLAGER et al) 15 February 2000 (15.02.2000), abstract.	1-18
A, P	US 6,004,267 A (TEWARI et al) 21 December 1999 (21.12.1999), abstract.	1-18
A	US 5,130,936 A (SHEPPARD et al) 14 July 1992 (14.07.1992), abstract.	1-18

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
29 June 2000 (29.06.2000)

Date of mailing of the international search report
11 SEP 2000

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
James P. Trammell

Telephone No. (703) 305-9769

Form PCT/ISA/210 (second sheet) (July 1998)